IDA

INSTITUTE FOR DEFENSE ANALYSES

The Effects of Automation on Battle Manager Workload and Performance

Amy Soller
John Morrison

January 2008

Approved for public release;
distribution is unlimited.

IDA Document D-3523
Log: H 08-000102

This work was conducted under contract DASW01-04-C-0003, Task CA-2-1617.01, for the JTAMDO. The publication of this IDA document of does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that Agency.

INSTITUTE FOR DEFENSE ANALYSES

IDA Document D-3523

The Effects of Automation on Battle Manager Workload and Performance

Amy Soller
John Morrison

PREFACE

This work was performed for the Joint Theater Air and Missile Defense Organization (JTAMDO) under the task "Analyses of Integrated Air and Missile Defense Battle Manager Tasks and Task Loadings in Support of Joint Theater Air and Missile Defense Operational Requirements and Architectures and Demonstrations." Technical cognizance for this task is assigned to CDR David Weller (JTAMDO). The Institute for Defense Analyses (IDA) point of contact (POC) is Dr. Kent Haspert.

ACKNOWLEDGMENTS

This work was sponsored by the Joint Theater Air and Missile Defense Organization (JTAMDO). The authors gratefully thank CDR David Weller, Mr. Jim Gill, and CDR Thad Dobbert for their support, interest, and understanding of the critical issues that directly affect the human warfighter.

The authors appreciated the direction, careful review, and recommendations provided by Dr. Kent Haspert throughout the development of this work. Dr. Frank Moses provided a comprehensive review of an earlier draft of this document and helped to shape and focus the content. Thanks also to Dr. Jack Laveson for contributing to the collection of reviewed literature.

Some of the literature reviewed for this document was requested and obtained directly from the Human Factors researchers who performed the work. All of these requests turned into illuminating discussions in which the researchers went out of their way to explain and clarify their work, enhancing the comprehensiveness and quality of this document. The authors are grateful for the contributions and collaboration of Dr. Barry Vaughan (U.S. Army Research Laboratory), Dr. Michael Paley (Aptima, Inc.), Mr. Dick Steinberg (Missile Defense National Team), Dr. Jared Freeman (Aptima, Inc.), and Mr. Sylvain Bruni (MIT).

Special thanks also to Mr John Everett (Technical Editor/IDA) and to the IDA Library staff, who helped us compile the resources for this review.

CONTENTS

Executive Summary ES-1

A. Introduction 1

B. Operator Performance Without Automation Assistance 3

1. Understanding Battle Manager Workload 3
2. Task Complexity 6
3. Battle Manager Cognitive and Behavioral Factors 10
4. Decision-Making Under Uncertainty in Stressful, High-Risk Situations 14
5. Cultural Differences 15

C. How Automation Can Augment Operator Performance 16

1. Varying Levels of Automation 17
2. Augmenting Air and Missile Defense Crew Performance Through Automation 24

D. How Automation Can Hinder Operator Performance 28

Automation Bias, Complacency, and Supervisory Control Effects 28

E. Summary 33

References Ref-1

Glossary GL-1

FIGURES

1. The complexity of the battlespace varies independently from the air traffic density 7
2. As the battle manager becomes overloaded, his performance will decrease. The point in time and rate of decrease will depend on his WM capacity 12
3. Without automation, operator performance is a function of workload, memory capacity, experience, and other operator-specific characteristics 13
4. Four stages at which automation can enhance system and human performance 18
5. Partial automation support for the Tomahawk missile-to-mission planners 22
6. Visual timeline decision aid for managing and scheduling the UAV ATO 32

TABLES

1. The metrics that will be used to quantify human and task-based factors that contribute to workload 5
2. A sampling of human-computer interface recommendations designed to alleviate ABL battle management workload 9
3. Levels of automation 18
4. Nine possible configurations that consider the battle manager's tasks and the three ABMA levels 19

EXECUTIVE SUMMARY

This report summarizes the literature reviewed in preparation for planning and executing a series of controlled, operator-in-the-loop (OITL) experiments to determine how an air and missile defense (AMD) battle manager's performance degrades with increased workload and how automated battle management aids (ABMA) can moderate this degradation. The sources for this survey range from studies that describe the basic limits of human memory capacity to those that assess the number of battle managers needed to operate a partially automated missile defense system.

The research indicates that without the assistance of automation, a battle manager's performance will degrade as the complexity of the task increases, in particular when he is tasked with attending to more than seven entities or decisions. Battle managers' performance may, however, vary considerably across experience levels and tasks. Prominent factors that affect the overall human-system performance include the battle manager's cognitive capacity and the system's level of automation.

This report outlines four different stages and eight different levels at which automation can enhance system and human performance. An abundance of research indicates that while automation may decrease operator workload, it may also decrease operator activity, engagement, and attention, which could lead to a decrease in situational awareness and performance. There is no shortage of research showing how overreliance on automation results in fatal accidents when the automated system fails.

THE EFFECTS OF AUTOMATION ON BATTLE MANAGER WORKLOAD AND PERFORMANCE

"People are flexible but inconsistent … machines are consistent but inflexible"

(Army Research Laboratory, 2005)

A. INTRODUCTION

As the U.S. air and missile defense (AMD) communities work toward a joint integrated AMD solution that includes a single integrated air picture (SIAP), integrated fire control, and automated battle management aids (ABMAs), the complexity of the system and time-sensitive threat environment will increase the cognitive demands placed on human battle managers. The future integrated environment will require cooperation among multiple Services and platforms (e.g., Patriot, Aegis, Theater High Altitude Area Defense (THAAD), fighter aircraft, Joint Land Attack Cruise Missile Defense (LACMD) Elevated Netted Sensor System (JLENS), Airborne Warning and Control System (AWACS), E-2)) to counter enemy air threats while ensuring the safe operation of friendly aircraft. As the number, variety, and capabilities of air and missile defense platforms increase, the ABMA's role in assisting the joint forces battle manager in tactical decision-making will become increasingly important. An ABMA that assists the battle manager by executing the best possible set of decision-making tasks at just the right time and to the most appropriate extent will optimize the overall human-system performance. To this end, the Institute for Defense Analyses (IDA) has been tasked to support the Joint Theater Air and Missile Defense Organization (JTAMDO) in determining the requirements for an ABMA. More specifically, IDA has agreed to design, plan, and execute a series of controlled, operator-in-the-loop (OITL) experiments to determine how an AMD battle manager's performance degrades with increased workload in the context of various realistic scenarios and how an ABMA can mitigate this degradation. This report summarizes the literature reviewed in preparation for planning and executing these OITL experiments. The insights from this literature survey and the planned experiments could be used to guide the development of a prototype ABMA. This prototype could then be deployed and studied in a realistic environment, such as the Virtual Warfare Center (VWC),[1] for verification and validation.

The objective of this literature review was to ensure that this research is novel and has not already been fully investigated, to validate the need to perform the planned experiments, and to gain insight into factors that may be important to consider while planning the OITL experiments and analyses. The OITL experiments will be designed to test the performance of battle managers under different workload constraints. These constraints will be simulated by increasing or decreasing the complexity of the experiment scenarios. The experimental design will also include conditions that alter the level of ABMA decision-support for each scenario (an example is outlined in Section C of this paper). The experimental participants are expected to be AMD battle managers who are responsible for managing (but not operating) theater missile batteries and groups of aircraft.

During the planned OITL experiments, AMD battle managers will be asked to make decisions in the context of realistic, simulated, theater-based threat scenarios. The level of difficulty and the level of decision support that the ABMA provides for each scenario will vary. The battle manager's performance will be assessed by measuring his ability to maintain situational awareness and make appropriate, effective, and timely battle management decisions. The decision process can be broken down into five steps, which are required to address a given threat scenario at a particular time: (1) identify the targets for sensor-shooter pairing, (2) establish the priority order for making pairing decisions, (3) determine sensor-shooter pairings, (4) assess which pairings meet the acceptability criteria, and (5) provide an ordered list of recommended pairings. The ABMA can take over any combination of these five steps; however, in all cases, the human battle manager must make the final sensor-shooter engagement decision.

This survey reviewed over 50 sources, primarily articles, books, and technical reports, related to the effects of automation on battle manager workload and performance in AMD-related domains. The sources range from those that describe the basic limits of human memory capacity to those that assess the number of battle managers needed to operate a partially automated missile defense system. They include studies within the AMD domain and across other similar domains, such as air traffic control. Because the

[1] The Boeing VWC is a multioperator, realistic, human-in-the-loop air and missile defense test bed. A typical VWC setup includes manned fighter aircraft, manned airborne sensors and surface-shooter platforms, and a sizable enemy raid consisting of a mix of missiles and manned aircraft.

term battle manager is specific to AMD-related domains, it will be used throughout this report to refer exclusively to an air and/or missile defense battle manager. Otherwise, the more generic and familiar term "operator" is used. Both terms refer to the human who is controlling or managing the system under discussion.

The following key questions were used to guide this review of the research and to organize the findings in this paper:

1. Without automation assistance, how many decisions can an operator handle per unit time? At what point does operator performance drop off, and does it drop off gradually or abruptly?
2. Under what circumstances will automation improve operator performance and optimize operator workload?
3. Under what circumstances might automation decrease operator performance and situational awareness while still optimizing operator workload?

None of the sources surveyed has fully investigated these questions in an AMD environment; however, many of the studies contain important implications for the planned OITL experiments. The next three sections summarize the literature that addresses each of the three questions. Section E summarizes some of the key historical findings.

B. OPERATOR PERFORMANCE WITHOUT AUTOMATION ASSISTANCE

This section reviews literature related to Question 1:

> *Without automation assistance, how many decisions can an operator handle per unit time? At what point does operator performance drop off, and does it drop off gradually or abruptly?*

The number of decisions that a battle manager can handle is largely determined by his workload. This section introduces and defines the concept of operator workload and then discusses the impact of task complexity and battle managers' behavioral factors on operator workload.

1. Understanding Battle Manager Workload

Operator workload (or simply "workload") is a human factor that describes the cognitive effort involved in performing a task. Understanding workload means understanding at what point, how, and to what degree the demands of the task or situation exceed the operator's available cognitive resources. Workload varies across operators and tasks. This variation stems from the number and complexity of the task(s), the ability, experience, and behavior of the operator, and the operational techniques, tactics, and procedures that are available and applicable. While consideration of these factors may seem straightforward, the interaction among them may produce nonlinear combinations of complex situations that result in uncertain outcomes. As such, similarly trained and experienced operators may respond differently to the same situations (Hilburn, 2004).

In the air traffic control domain, operator workload has traditionally been measured using subjective assessment instruments such as the National Aeronautics and Space Administration (NASA) Task Load Index (TLX) (Hart & Staveland, 1988), the Subjective Workload Assessment Technique (SWAT) (Reid & Nygren, 1988), and the Workload Profile (Tsang & Velazquez, 1996). Rubio, Díaz, Martín, & Puente (2004) compare these three subjective assessment instruments, all of which involve asking the operator to self-assess his workload by considering factors such as stress level, effort, and mental demand.

In analyzing the data from our planned OITL experiments, we will take a different approach. We will calculate workload objectively as a combination of contributions not only from subjective workload measurement instruments, but also from objective performance metrics and task complexity. In particular, workload will be measured post-hoc by computationally estimating the quantity and complexity of decisions that the AMD battle manager has actually made for a given scenario (at a particular time, *t*). It can only be calculated after the battle manager has completed the task (up to time *t*).

The workload analysis for the planned OITL experiments will consider not only subjective workload measurement instruments, but also objective performance metrics and task complexity.

Table 1 shows all the factors that will be considered and their corresponding assessment metrics. The human factors are listed in the top half of the table, while the task-based factors are listed in the lower half of the table. Task-based factors are largely derived from the complexity of the scenario. Task complexity (described in the next section) will be calculated by considering the Inherent Task Complexity and the Actual Task Complexity (the last two rows of Table 1). The Inherent Task Complexity is derived from the scenario (see Section B.2 for the theoretical foundation and examples of this concept). It is strictly a task-related factor—independent of operator performance—that

Table 1. The metrics that will be used to quantify human and task-based factors that contribute to workload

	Factor	Metric
Human Factors	Experience	Demographic questionnaire
	Stress level	NASA TLX SWAT Assessment Observer reports
	Confidence	Logged performance data SWAT Assessment
	Attention	Observer reports SWAT Assessment
	Individual differences	Strategies and skills applied during the scenario and gathered from logged performance data, observer reports, and After Action Reviews (AARs) Demographic questionnaire
	Performance	Logged performance data # of shots fired/missiles launched # of hits/# of misses, # of leakers, # of impacts Effectiveness Speed Efficiency Commonality
Task-based Factors	Scenario: Inherent Task Complexity	Raid size Blue force laydown Red force laydown Routing Defended assets Timing Order of events
	Actual Task Complexity	A measure of the Inherent Task Complexity reduced by the number and complexity of the tasks performed by the ABMA

describes the complexity of the scenario at any given time. It accounts for the number of decisions involved in the scenario and the difficulty (or complexity) of each decision. Both static and dynamic scenario-based elements contribute to the inherent task complexity in AMD. The ordering of scenario events also affects the inherent task complexity (Leonard Adelman, Bresnick, Christian, Gualtieri, & Minionis, 1997).

The metrics, shown on the right-hand side of the table, include both standard assessment instruments (described earlier in this section) and variables that would typically be computed and logged in an AMD simulation. For example, variables that affect the Inherent and the Actual Task Complexity include the raid size, blue force laydown, routing, defended assets, and timing. None of these variables change across operators.

The Actual Task Complexity is the Inherent Task Complexity reduced by the number and complexity of the tasks performed by the ABMA. If no ABMA is present, the Actual and Inherent Task Complexity values are the same.

By combining the human and task-based factors in Table 1, estimates of workload can be calculated per unit time, accounting for changes in task complexity and other factors that change across a scenario. This notion of workload accounts for the decision density (after-action recounting of the number of decisions the operator made per unit time) and the degree of difficulty of each decision for each operator. The next four sections (B.2–B.5) explain why the literature suggests that different operators will experience different workloads for the same scenario and how factors such as the operators' level of experience, stress, confidence, and other human factors will influence their performance.

2. Task Complexity

In general, an operator's performance is expected to degrade as the complexity of the task increases. Task complexity is not the same as air traffic density (in air traffic control) or air raid density (in missile defense). Through the late 1960s and 1970s, research suggested that air traffic density and radio communications were the main contributors to air traffic controllers' workload (Hurst & Rose, 1978; Mogford, Murphy, & Guttman, 1994). Although these factors do contribute to workload, Mogford et al. (1994) showed that factors such as the relative frequency of complex as opposed to direct aircraft routings and the need for arrival/departure sequencing and spacing may be more significant.[2] Other similar studies suggest that factors such as the mixture of aircraft types, the climbing and descending of aircraft flight paths (Histon & Hansman, 2002; Mogford, Guttman, Morrow, & Kopardekar, 1995), and the degree to which the structure of the airspace dynamically changes (Cummings & Tsonis, 2006) also contribute significantly to task complexity.

A battle manager's performance is expected to degrade as the complexity of the task increases. Complexity factors include the timing, the quantity and order of events, and the degree of uncertainty.

2 This study involved administering a sequence of questionnaires to over 50 air traffic controllers at the Federal Aviation Administration's Jacksonville Air Route Traffic Control Center.

Hilburn's (2004) study includes a meta-review and analysis of 25 other studies that identify factors that contribute to task complexity in air traffic control.

Figure 1 illustrates how the complexity of the situation may vary independently from traffic density (Hilburn, 2004). Just three aircraft have changed their orientations from the figure on the left-hand side to the one on the right-hand side. The air traffic density is unchanged, yet the complexity of the situation has increased dramatically.

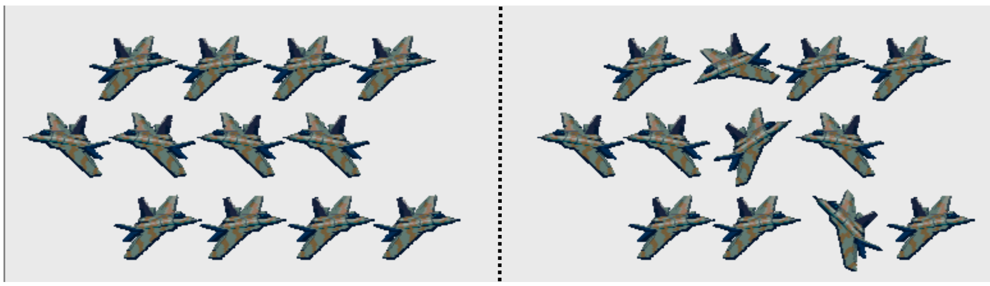

Figure 1. The complexity of the battlespace varies independently from the air traffic density (Hilburn, 2004)

The timing and order in which events occur also affect the complexity of the task. The speed at which events occur in the scenario affects the number of tasks the operator must attend to over a bounded period of time (Cannon-Bowers & Salas, 1998). Events containing uncertain information further increase the complexity. In the Adelman, Bresnick, Christian, Gualtieri, & Minionis (1997) study, 43 Patriot air defense operators were asked to identify simulated incoming aircraft as friendly or hostile and then engage those that were determined to be hostile. As aircraft entered the airspace that contained protected assets, the participants were given sequences of conflicting information regarding the interpretation of the aircraft in question (e.g., the aircraft responded as a Friendly to an Interrogation-Friend-Foe inquiry and then jammed the Patriot's radar). When the information in these sequences was reordered slightly, the operators' judgments about the unknown aircraft changed significantly. This example shows how changing the order in which information is presented can affect the complexity of the task.

Task complexity can be mediated by the fidelity and the design of the operator display. The display is composed of static and dynamic environmental constituents, each of which contributes to the overall complexity. Static elements include geographic variables such as terrain, land and sea boundaries, situated sensors, and other assets that do not change. Dynamic elements include moving aircraft, missiles, and weather conditions (if they change during the time period under consideration). The depiction of the static and dynamic elements on the interface can also affect the complexity of the task. For example, the interface may affect an operator's ability to communicate and issue critical commands, to locate "hot spots" (locations where critical events often occur), to manage potential conflicts, and to visualize groups of aircraft or other entities as generalized representations that can more easily be managed (Histon & Hansman, 2002).

McDermott, Klein, Thordsen, Ransom, & Paley (2000) provide an illustrative example of how battle manager display features can be modified according to the complexity of various tasks. They conducted cognitive task analysis interviews with Airborne Laser (ABL) program managers (PMs), an ABL subject matter expert (SME), and several crew members who had participated in the Joint Expeditionary Force Exercise 1999 (JEFX-99). From these interviews, the researchers developed a sorted list of tasks, decisions, and functions involved in ABL missions. The relative time to complete each task and the relative workload each task imposed on the ABL battle manager were also estimated. Each task was also assigned a rating (high, medium, or low) to indicate how cognitively challenging it was. In our terms, this rating represents the inherent task complexity. For some of the more cognitively complex tasks, McDermott et al. provided interface modification recommendations that were observed to alleviate the ABL battle manager's workload. These recommendations included options such as enabling the battle manager to toggle track numbers on or off, to put their own designators on tracks, and to calculate the range between objects. Table 2 includes a sampling of recommendations relevant to AMD and the battle management function that they were intended to address in the ABL domain.

Later in this paper, Sections C and D explain how the level of automation, or in our case, the ABMA, acts as another factor that can mediate the complexity of the task and possibly reduce the operator's perceived workload. The ABMA is not likely to affect operator performance in a linear fashion. The degree to which the ABMA will affect a battle manager's performance will depend on other factors, such as his domain experience and his ability to adapt to the ABMA. For example, the utility of the ABMA will likely increase as the inherent task complexity increases. As the scenario becomes more complex, the operator will eventually become overloaded and will need to rely on the ABMA. However, as the operator begins to rely more on the ABMA, he may perceive that the complexity and difficulty of the task decreases, perhaps significantly. The operator may gradually begin to play a monitoring role rather than that of an active controller role. This means that the task complexity, as it is perceived by the operator,

Table 2. A sampling of human-computer interface recommendations designed to alleviate ABL battle management workload (from McDermott et al., 2000)

Display Concept Recommendation	ABL Battle Management Function Addressed
Include designators for tracks	Monitor enemy tracks
Include range rings for surface-to-air missile sites	Gauge threat to ABL
Enable operators to differentiate between track types (e.g. surface, air, ground) and toggle track numbers on or off	Filter and sort information
Allow operators to put their own designators on tracks	Detect problems and inconsistencies in track data
Make high-value assets salient on the display	Monitor location of high-value assets
Create an automated Air Tasking Order (ATO) that battle managers can access from their displays	Reassign tasks/orchestrate priorities
Record information about missile launches (e.g., launch time and location, track number, actions taken, results)	Determine trends of launch locations
Use the information about past missile launches to predict and prepare for future launches	Anticipate future launches
Enable the battle manager to calculate the distance between objects	Recommend changes in orbit or speed
Display messages to inform the operator why the system cannot execute an instruction (e.g., system cannot fire because of inability to acquire target)	Know weapon status and if weapon is ready to fire
Show two correlated displays from the same perspective	Deconflict missiles from different locations
Sound an audio alarm and dim all other tracks when a missile is launched so that the new threat is easily detected	Detect and recognize launch
Allow the battle manager to zoom in and out to get a better picture	Report results

varies with the level of automation set by the ABMA. From an analytical perspective, this renders the task complexity inappropriate as an independent variable. Section D addresses this in more detail.

3. Battle Manager Cognitive and Behavioral Factors

Task complexity is just one factor that contributes to an operator's workload. The operator's performance will also depend on his experience and cognitive ability (Mogford et al., 1995). As early as 1955, experimental evidence revealed that the amount of information a person can cognitively process at once does not increase linearly with the amount of information presented to him (G. Miller, 1955). A well-known limit to memory capacity is present in all domains (about 7 items), and this limit applies even across fundamentally different stimuli. This short-term memory capacity limitation manifests itself in a variety of tasks and materials, but, most typically, it is measured by memory span tasks. In these tasks, subjects are presented several unrelated items at a standard rate and asked to recall them in order. Memory span is defined as the maximum number of items that can be recalled correctly.

As originally conceived, short-term memory capacity is a fundamental human capability that underlies a variety of cognitive tasks. In reality, performance on memory span tasks correlates with performance on similar rote memory tasks, but it does not necessarily relate to performance on more complex tasks that would seem to depend on short-term memory, such as reading comprehension.

Almost 20 years after the introduction of Miller's concept of short-term memory capacity, Baddeley & Hitch (1974) introduced the notion of working memory (WM). The WM concept viewed short-term memory capacity as the result of a dynamic executive that controls temporary storage, rehearsal, and attention processes. In a simplistic sense, the WM concept is more inclusive than short-term memory because it accounts not only for short-term storage, but also for all the processes that control it.

As the WM concept matured, researchers began to develop ways to measure it. The breakthrough was the realization that if WM includes the process for allocating attention, a WM test must require the performer to cope with multiple memory demands. Thus, WM is measured using the "dual-task" paradigm wherein a person is asked to do two or more qualitatively different tasks simultaneously. These dual tasks take on a variety of forms, but the task described by Engle (2002) is representative:

> … Subjects read aloud a series of operation-word strings such as 'Is 4/2 + 3 = 6? (yes or no) DOG.' They respond as to whether or not the equation is correct then read the capitalized word aloud. After a set of two to seven such operation-word strings, we measure the number of words recalled …

In contrast to memory span performance, WM task performance correlates with a wide range of higher order cognitive tasks, such as reading and listening comprehension, the ability to follow directions, note taking, reasoning, bridge playing, and even writing computer programs. Engle speculates that WM performance, especially WM with a simple storage capacity statistically controlled, corresponds to the fluid intelligence construct. Fluid intelligence is the ability to draw inferences and relationships in new problems, independent of acquired knowledge.

The WM concept complicates the answers to JTAMDO questions, such as "how many unassisted decisions per unit time?" Engle (2002) discussed how the concept of WM changes our notions of short-term memory capacity, which

> … often conjures up images of a limited number of items or chunks that can be stored (e.g., 7 ± 2). However, my sense is that WM capacity is not about individual differences in how many items can be stored per se but about differences in the ability to control attention to maintain information in an active, quickly retrievable state … (p. 20).

On the other hand, compared to the traditional static concept of short-term memory, the concept of WM seems more relevant to performance on command and control (C2) tasks, particularly the time-sharing demands of such activities. For example, Adelman, Miller, and Yeo (2004) showed how an operator's WM capacity can directly affect his performance in air defense tasks. Their task involved determining the threat level of air-breathing targets that enter a set of concentric rings on a radar display and making engagement decisions for those targets. Participants were given the airspeed, course, and range information for each target, and, in some experimental cases, they received additional altitude and radar information. The rate at which targets appeared was also manipulated to vary the time pressure across experimental conditions. Before performing the task, the participants completed a WM capacity test in which they viewed numbers that flashed in sequence on a monitor and determined whether each number was the same as the one that flashed one, two, or three numbers earlier in the sequence. Adelman et al. (2004) found that performance on this WM task correlated positively with participants' decision accuracy on the air defense task. The largest effect occurred in those situations in which participants were asked to consider the maximum quantity of information (including the additional altitude and radar information) to make a decision.

An operator's WM capacity can directly affect his performance in air defense tasks.

If human memory capacity is limited, as the psychological and air defense research suggests, we should expect the performance of an AMD battle manager to decline rapidly when he is overloaded with more than seven entities or decisions. For example, if the level of automation (i.e., ABMA) is held constant or turned off and the complexity of the scenario (e.g., number of threats per unit time) is gradually increased, the operator's effectiveness and efficiency should begin to decrease at some point in time (see Figure 2).

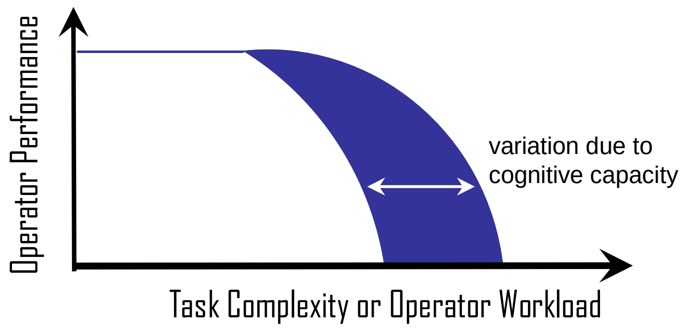

Figure 2. As the battle manager becomes overloaded, his performance will decrease. The point in time and rate of decrease will depend on his WM capacity

A person's experience in a domain can also change his cognitive and mental capacity with respect to that domain. Chess is one of the complex cognitive domains that has been studied in great depth to assess the degree to which experience affects mental capacity. Simon (1974) describes the performance of novices and grandmasters who were asked to reproduce chess board configurations after they were given 5 to 10 seconds to study them:

> If the pieces represent a position from an actual game (unknown to the subjects), then grandmasters and masters will generally reproduce the position (about 20 to 25 pieces) almost without error, while ordinary players will generally be able to place only a half dozen pieces correctly. If the same number of pieces is placed on the board in a random pattern, grandmasters and ordinary players alike will be able to place only a half-dozen pieces correctly (p. 487).

This effect has been replicated in other, more complex domains such as physics (Larkin, McDermott, Simon, & Simon, 1980), electronics troubleshooting (Gott & Lesgold, 2000), and air traffic control (Mogford et. al, 1995). It can be explained in part by chunking: mentally recoding items into aggregates that can more easily be recalled and cognitively processed. Extensive practice in a domain (Newell & Rosenbloom, 1981) is likely to result in highly efficient methods for chunking and applying mnemonics and other domain-specific memory recoding schemes, thus enabling a person to become highly efficient in recognizing and remembering domain-specific elements, procedures, or situations. Over the past 30 years, the cognitive science community has learned that the development of expertise involves much more than improved access to items in memory. It renders significant changes in performance and process. Experts recognize recurring patterns and act on compiled combinations of principles and procedures rather than serially and systematically considering and processing individual pieces of information (Anderson, 1982; Larkin et al., 1980). They follow cognitive procedures that they have automated through knowledge and practice and become efficient in restructuring their own knowledge to select and evaluate alternatives when necessary (Gott & Lesgold, 2000).

In our case, the degree of error across a pool of operators who have different characteristics is likely to vary significantly based on backgrounds and experience. It will depend on the quantity, complexity, and context of the decisions that the operator makes at each point in time during a given scenario. It can be calculated per unit time to account for changes in the varying complexity of the task and other factors that change across a scenario. Figure 3 shows how these factors might be considered for characteristic groups of operators (e.g., novice, intermediate, expert). The most experienced operators might have performance curves in the yellow area, indicating that they are able to handle a more difficult, complex scenario without as much performance degradation. The point and the rate of performance degradation (slope of the line) are unlikely to converge across diverse communities of operators. These factors will be influenced by the operator's personal characteristics, including his mental capacity, experience, and confidence.

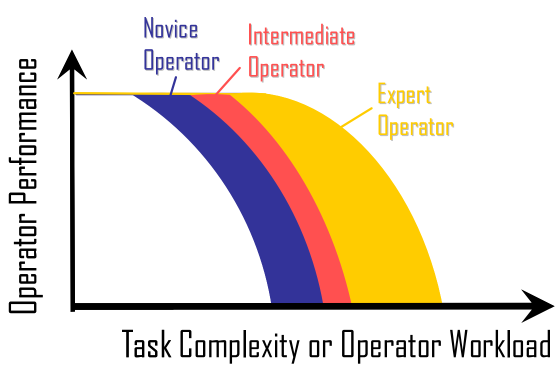

Figure 3. Without automation, operator performance is a function of workload, memory capacity, experience, and other operator-specific characteristics

The term "workload" is context specific and describes the after-action recounting of the number of decisions the operator made per unit time combined with the subjectively weighted degree of difficulty of each decision. Different operators will experience different workloads at different times for the same scenario. Factors such as the operators' level of experience, stress, confidence, mental capacity, and other human factors will influence their workload, which, in turn, will influence their performance.

Different operators will experience different workloads at different times for the same scenario.

4. Decision-Making Under Uncertainty in Stressful, High-Risk Situations

A battle manager's decision-making behavior is affected by the way he handles uncertainty in stressful, high-risk situations. Because decision-making in an AMD environment involves a high degree of risk, it is worthwhile to consider how a battle manager's decision process might change when he is presented with risky situations involving uncertain information. Research on human decision-making (Kahneman & Tversky, 1979) has shown that the way people perceive risk and exhibit risk-seeking or risk-aversion behavior cannot be explained in a computationally logical way by expected utility theory. Kahneman & Tversky (1979) developed Prospect theory to explain human decision-making behaviors in the presence of risk. The theory's underlying concept is that people base their judgments on perceived increases or decreases in value caused by gains or losses (with respect to some reference point), with less regard for the final outcome. The theory also states that as a person accumulates losses without adapting his reference point, his tendency toward risk-taking behavior increases (which explains the tendency of some gamblers to increase their betting during a losing streak). This research may be important to consider if the human battle manager, who may be applying Prospect theory to make decisions, misunderstands the ABMA's activities because it is applying a logical utility-based theory to make decisions on behalf of the battle manager.

According to Prospect theory, when people are presented with the possibility of winning, they tend to select choices that minimize risk and maximize certainty, even when the risk is insignificant. The one exception seems to be situations in which all the choices present similar gains and losses and "winning is possible, but not probable" (p. 267). A good example of this situation is the $5 lottery ticket for which there is a very small chance of winning a large sum of money. In that case, people tend to select the choice that offers the greatest possible gains and accept the small possible loss. On the other hand, when people are presented with the possibility of losing (instead of winning), they exhibit risk-seeking behavior that attempts to minimize loss, at the risk of losing even more.

The way that a decision-making problem is presented can significantly affect the resulting decision. For example, a problem in which the decision-maker is given $1,000 and asked if he would like an additional $500 (the risk-averse decision) or a 50% chance to win another $1,000 (the risk-seeking decision) can also be presented as one in which he is given $2,000 and asked if he would like to give back $500 (the risk-averse decision) or risk a 50% possibility of losing $1,000 (the risk-seeking decision). In the first case, the decision-maker is likely to make the risk-averse decision. In the second case, he is likely to take the risk-seeking choice. When problems are broken down into subproblems that are presented sequentially, each requiring an independent decision, the final outcomes may also differ.

Kahneman and Tversky's (1979) Prospect theory may have far-reaching implications in AMD. Once a battle manager has attained an understanding of the current battle situation, Prospect theory indicates that changes in the situation are more likely to affect his decision than his consideration of the decision outcome. If he perceives that the situation is changing to favor friendly forces, he may choose a risk-averse decision to minimize risk and maximize certainty; however, if he perceives that the situation is changing for the worse, he may decide to make riskier choices. Whether this type of innate human behavior is representative of AMD battle managers' decision-making processes and can be mediated through training or an ABMA remains an open question.

5. Cultural Differences

Cultural differences may also affect a battle manager's decision-making behavior. In a study funded by the U.S. Air Force Research Laboratory (AFRL) from 2001–2004, Micro Analysis and Design, Inc. (MA&D)[3] assessed the contribution of cultural factors to operator performance in an Integrated Air Defense System (IADS) (Mui et al., 2004). The cultural factors that they considered—distribution of power, willingness to take risk, and familiarity with the enemy—were derived from an analysis of differences among cultures performed by Hofstede (1984) and consequently described by Klein, Pongonis, and Klein's (2000) Cultural Lens model. Although MA&D was able to model a range of

---

[3] MA&D was acquired by Alion Science and Technology in 2006.

values for each cultural variable, the ranges for the IADS study were simplified to "high" or "low."

Mui et al. (2004) studied three scenarios and two countries of interest (Iraq and North Korea). The first scenario was a prewar scenario in which two enemy F-16s patrolled a nearby border. The second was a traditional wartime scenario in which the blue forces were tasked to defend an area from 69 invading aircraft. The third was an unconventional scenario incorporating surprise attacks.

The MA&D study is similar to our planned OITL experiment in that the simulated IADS operators were working in the Sector Operations Center, were making critical high-level command decisions, and did not have direct control over the early warning (EW) radars or weapons. Values for the cultural variables were determined from interviews with SMEs. North Korea was assigned a high willingness to take risk, and Iraq was assigned a low willingness to take risk. These cultural factors affected the outcome of the scenario in relatively predictable ways. For example, a country's willingness to take risk translated into more firings on unknown aircraft. In the first scenario, North Korea was much more likely than Iraq to acquire an unknown aircraft with targeting radar to persuade the aircraft to retreat. Likewise, assigning a country a low familiarity with the enemy translated into less effective and less successful offensive strategies. Although this simulation was somewhat contrived, it did show how cultural factors can affect the order, time, and locations of firing assignments. It is not in our current plan to consider cultural variables in our OITL battle manager study. Instead, individual differences that account for performance-related differences across all cultures will be taken into consideration (see "individual differences" in Table 1).

C. HOW AUTOMATION CAN AUGMENT OPERATOR PERFORMANCE

This section reviews literature related to Question 2:

Under what circumstances will automation improve operator performance and optimize operator workload?

Automated systems have the potential to increase human performance by carrying out certain mundane functions, allowing the human to concentrate on more complex cognitive tasks. For example, the cruise control system on a vehicle alleviates the need for the driver to regulate his speed and allows him to concentrate on other vehicles' motion, on street signs, and so forth. Automated systems can also augment human activity by carrying out tasks that humans are not physically capable of performing (e.g., weather satellite imaging) or by performing tasks for which humans show inherent limitations (e.g., real-time calculation of the distance to a target).

1. Varying Levels of Automation

Battle managers' performance will vary depending on the specific tasks automated by the ABMA. Not all tasks, however, are good candidates for automation. Kaempf, Wolf, and Miller (1993) studied the decision-making processes of an air-to-air warfare team in the Combat Information Center of Aegis cruisers. The researchers found that the most difficult operator tasks involved assessing the situation and obtaining the information needed to maintain good situational awareness (as opposed to deciding whether to engage a threat). By the time the operators were ready to make the decision to engage, they had already obtained the information they needed. At that point, they just followed the instructions set out in the standard operating procedures (SOPs). This work indicates that the ABMA will affect operator performance if it assists the operator with the most complex decision-making tasks, including assessing the situation.

Automation can enhance system and human performance at four different stages (see Figure 4) (Sheridan & Parasuraman, 2006):

- The first stage involves the acquisition of information (e.g., from sensors or fire units via communication networks).
- The second stage involves the representation and display of the information on the human-machine interface (HMI). Although automation during this stage is not necessarily aimed at decision support, it can make a significant difference in performance. For example, a study by Smith, Johnston, and Paris (2004) showed that Naval officers in an Aegis Combat Information Center who viewed information on set of specialized displays were significantly less likely to misclassify and target commercial aircraft than the Naval officers who used a standard Navy training system.
- The third stage at which automation can enhance performance is the decision-making stage. Our planned OITL experiment includes provisions for two different experimental baselines. The first baseline represents the current system state in which none of the stages are augmented through automation. For example, under this baseline, the battle manager must request status information from sensor and fire units. The second baseline includes system functions that automate the first two stages (information acquisition and display). Building upon this second baseline, our planned OITL experiment will be designed to test three distinct types of automated decision-aiding that augment the third stage (decision-making).

- The fourth stage is the implementation stage; however, this will not be addressed in this study because of the focus on battle management, command, and control.

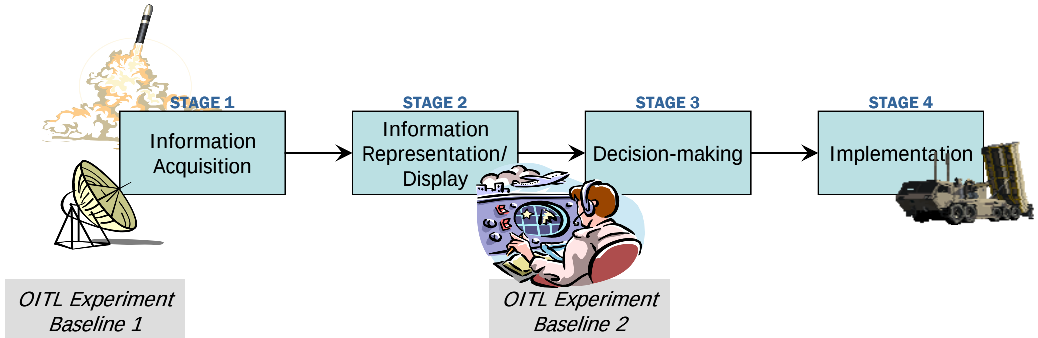

Figure 4. Four stages at which automation can enhance system and human performance (adapted from Sheridan & Parasuraman, 2006)

The planned OITL experiments will focus on identifying the type of automation that augments the third stage in Figure 4 (decision-making). The type of automation varies along two dimensions: the specific decision-making task being automated and the level at which the system is automating that task. Table 3 shows the eight levels of automation that have been documented and applied in practice (also see Sheridan, 1992; Sheridan & Parasuraman, 2006). Each level can be applied at each of the four stages described previously.

Table 3. Levels of automation (from Sheridan & Parasuraman, 2006, p. 94)[4]

Level	Response
1	The computer offers no assistance. The humans must do it all.
2	The computer suggests alternative ways to do the task.
3	The computer selects one way to do the task and
4	executes that suggestion if the human approves OR
5	allows the human a certain amount of time to veto before automatic execution OR
6	executes the suggestion automatically and then necessarily informs the human OR
7	executes the suggestion automatically and then informs the human only if asked.
8	The computer selects the method, executes the task, and ignores the human

4 This table is a condensed version of the full 10-level taxonomy described in Sheridan (1992; Sheridan & Verplank, 1978).

Level 1 in Table 3 represents our second baseline for the planned OITL experiments. Information regarding the status of sensor and weapon systems and inventories are automatically acquired and displayed on the battle manager's interface. The human operator must then perform the tasks of (1) assessing the situation and prioritizing threats, (2) pairing weapons to targets, and (3) determining the acceptability of the engagement selections. For each of these tasks, the ABMA has the option of suggesting alternative ways to do the task (Level 2), selecting one way to do the task but allowing the human operator to override this selection (Level 5), or performing the task (Level 8). This results in nine possible ABMA configurations, in addition to the two baselines (see Table 4).

Table 4. Nine possible configurations that consider the battle manager's tasks and the three ABMA levels

	ABMA Levels		
Battle Manager Task	**ABMA suggests alternatives (Level 2)**	**ABMA selects one way to do the task, allows manual override (Level 5)**	**ABMA performs the task (Level 8)**
Assess and prioritize threats	ABMA suggests possible threat prioritizations	ABMA selects threat prioritization, allows manual prioritization changes	ABMA prioritizes threats automatically
Pair weapons to targets	ABMA suggests all possible weapon-target pairings	ABMA selects weapon-target pairings, allows manual pairing changes	ABMA pairs weapons-to-targets automatically
Determine the acceptability of the engagement selections	ABMA suggests acceptable engagement selections	ABMA selects acceptable engagement, allows manual engagement overrides	ABMA determines acceptability of engagements and performs engagement

The experimental options listed in Table 4 will allow us to determine which levels of ABMA improve the human information processing and decision-making capabilities for each of the three battle manager tasks. While our planned OITL experiments appear be novel to AMD, similar studies have been executed in other related domains, including ballistic missile defense (BMD), Tomahawk strike planning, and air traffic control (described in that order in this section).

In 2005, the Schafer Corporation (Schafer Corporation, 15 January 2005) studied the effect of the first two levels of automation in Table 3. They developed an automated decision aid in the form of an intelligent agent for the Ground-Based Midcourse Defense (GMD) Fire Control (GFC) system. This project was funded on a Phase II Small

Business Innovative Research (SBIR) contract sponsored by the GMD program management office (GFC Products Division). The Northrop Grumman Corporation, which was in the process of developing the GFC, subcontracted to Schafer to develop and test the decision aids. During the testing, GFC battle managers were asked to decide whether and when to override the automated battle management algorithms. Such decisions included engaging a track, ordering a cease fire, and setting the minimum and maximum number of intercepts allocated to negate the reentry vehicle (RV) or track cluster. The decision aid performed four primary tasks, all of which involved highlighting portions of the display to raise the battle manager's awareness of certain conditions. The highlighting was applied to (1) asset values, (2) RV likelihood information in cases with uncertainties due to sensor tracking, (3) RV likelihood information in cases with uncertainties due to booster parenting, and (4) clusters of tracks that share the same impact region or booster parent as another missile with an override. Schafer Corporation conducted an experiment that tested the performance of 15 battle managers, with and without the decision aid. The participants included uniformed GFC battle managers and civilian SMEs. Their performance was measured by calculating their task accuracy and reaction time. In all the conditions, the battle managers completed the tasks faster with the decision aid, and, in two of the four conditions, they completed the tasks significantly faster. In all the conditions, the battle managers' accuracy was also significantly better with the decision aid.

Decision aids that raise battle managers' awareness of critical conditions can increase task accuracy while decreasing latency.

Cummings and Bruni (in press) studied the effect of the first three levels of automation in Table 3 on Naval operators' decision-making performance in a Tomahawk Land Attack Missile (TLAM) planning domain. The study involved the development of automated and partially automated decision aids that assist Tomahawk strike planners in a multiple resource allocation problem. The planners were asked to assign missiles to missions by taking into account factors such as the characteristics of each of the planned missions (e.g., target, route, launch basket), the characteristics of the available missiles (e.g., type, ship and launch basket required, warhead), each ship's rate of success for missile launches, and other constraints such as the number of days to port for each candidate ship.

The first interface Cummings and Bruni (in press) tested (Interface 1) required the operators to perform the missile-to-mission matching manually. The interface did filter the available information, which prevented the operators from matching missiles to missions in unfeasible combinations. The second interface (Interface 2, shown in Figure 5) provided some decision-support tools. These tools included tables showing missiles that can be matched to missions according to criteria, such as priorities, that the operator can enter. It also included an “Automatch” button that automatically matched and prioritized missiles to missions in order of mission importance. This version allowed the operator to perform “what if” comparisons and save them for future planning. The third interface (Interface 3) was a higher level display that did not graphically represent specific missile-to-mission pairings. It required the user to input his constraints, criteria, and priorities using graphical slider bars. The automated system then attempted to optimize the resources available to meet the given criteria and produced the best possible missile-to-mission matches according to an optimization algorithm.

Twenty U.S. Naval officers tested five combinations of the interface designs: Interfaces 1, 2, and 3 separately, Interfaces 1 and 3 together, and Interfaces 2 and 3 together (Bruni & Cummings, 2007). Operator performance was measured by an objective weighting function that calculated a weighted sum of the percentages of correct missile-to-mission matches according to mission priority. Those missions that had higher priorities contributed more heavily to an operator’s measure of performance. The results showed that Interface 1, the manual matching interface, and the combination of the two automated decision-support interfaces (Interfaces 2 and 3 together) generated significantly better operator performance than the three other conditions. Interface 1 may have produced good results because the operators explained that they were familiar with similar types of manual missile-to-mission matching interfaces; however, this explanation does not indicate why the combination of Interfaces 1 and 3 generated the worst performance.

Cummings and Bruni (in press) also found that “the highest level of automation, Automatch, seemed to improve the mission-missile matching process. According to users’ feedback, the Automatch function allowed for faster computation of solutions. However, it was not always used, and, in many cases, participants exhibited significant distrust in the Automatch, by constantly cross-checking the automation’s solution, which was expensive in terms of time” (p. 12).

Automated decision aids will assist the human operator as long as he can still access the decision-making data.

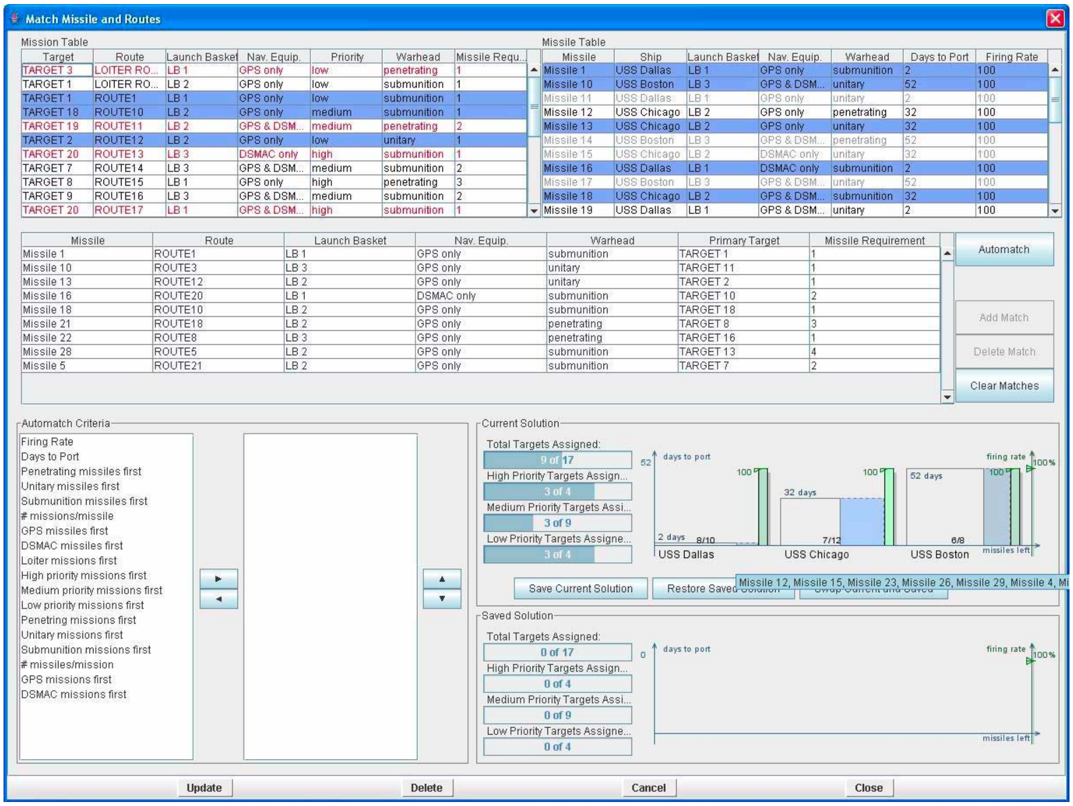

Figure 5. Partial automation support for the Tomahawk missile-to-mission planners (Cummings & Bruni, in press)

The third interface (which required users to manipulate graphical slider bars to denote constraints and priorities and automatically performed the matching) induced user frustration and mistrust because users "did not have access to the raw data and did not know exactly what assignments were made or if a specific missile was available. This inability to 'drill down into the detail' is a known limitation of configural displays; however, participants were able to adjust their strategies accordingly and performed as well as participants with other interfaces" (Cummings & Bruni, in press, p. 13). Participants who used Interfaces 1 and 2 explained that they felt compelled to look at all the available drill-down information, even if it was not significant, to ensure that they did not miss any critical information. This behavior led to an increase in the solution time (Bruni & Cummings, 2007).

This study and findings demonstrate the utility of automation in assisting the human operator and the importance of ensuring that the operator can access the most essential data and pairing options—even when the automated system will be recommending or making pairing decisions.

The GMD study done by the Schafer Corporation (15 January 2005) and the TLAM study done by Bruni and Cummings (2007) demonstrate the value of automated decision aids that operate at Levels 2, 3, and 4 in Table 3. These tools suggest alternative ways to do the task, recommend a particular way to do the task, and may even execute the recommendation if approved by the human operator. The level of automation chosen for implementation will affect the overall human-system performance. This issue is addressed next.

In the late 1990s, the National Academy of Sciences (NAS) convened a panel on Human Factors in Air Traffic Control to determine what levels of automation are most appropriate for which air traffic control tasks (Wickens, Mavor, Parasuraman, & McGee, 1998). One of the panel's main findings was that decision aids should not go beyond suggesting preferred alternatives in situations that involve a considerable degree of uncertainty and risk. The reasons for this caution include loss of situational awareness, complacency, and skill degradation and are described later in Section D. Decisions about which tasks to automate and to what degree should consider the reliability[5]—not the availability—of the automation (Hawley & Mares, 2006). The panel also recommended that the choice of automation level should be based on an understanding of human behavioral strengths, tendencies and vulnerabilities, and the consequences of making mistakes.

Decisions about which tasks to automate and to what degree should consider the reliability—not the availability—of the automation.

All the NAS panel findings are directly applicable to AMD and will be integrated into our study. The first stage of the planned OITL study (which will address Question 1) is designed to foster an understanding of the cognitive and mental capacity of AMD operators who have varying levels of experience. Then, the performance of each group of operators (e.g., novice, intermediate, expert) will be tested for each battle manager task and ABMA level. Both the reliability of the ABMA and the resulting improvement in performance will be considered in assessing the risks and benefits of automating each task.

The choice of the automation level should be based on an understanding of human behavioral strengths, tendencies and vulnerabilities, and the consequences of making decisions.

---

5 Reliability refers to the decision-making competence of the automation, not to its operational consistency.

2. Augmenting Air and Missile Defense Crew Performance Through Automation

The benefits of automated systems go beyond enhancing individual human performance, to changing environmental and crew configuration requirements. Although our planned OITL experiments will consider the performance of an individual battle manager under varying levels of automation, battle managers seldom work alone. They operate as part of a crew and have roles such as weapons assignment officers, air defense managers, communications officers, and battle manager chiefs. This section addresses considerations in constructing and sustaining crews of operators and automated systems with complementary responsibilities. Many of the findings from these crew performance analyses substantiate what we already know about individual battle manager performance. They also illuminate the key concepts that will need to be remembered in the future when the individual OITL experiments are expanded to crew-based battle management experiments.

Increasing the number of crew members causes an exponential increase in the number of ways that variables such as crew members' behavioral factors, cognitive abilities, and experience can be combined with the candidate system functions and their corresponding levels of automation. In these situations, identifying the most appropriate combinations of human and system functions to operate in the environment becomes difficult. An unreasonable number of traditional human-in-the-loop experiments would need to be designed, executed, analyzed, and synthesized to account for all possible experimental conditions. Computer-based Human Behavior Representations (HBRs) in constructive simulations present an alternative because they do not require human operators. HBRs have been used successfully to simulate human behaviors, cognition, and performance in complex military environments (Morrison, 2003). They can become complex; for example, some include models of short-term memory, long-term memory, and emotional behavior. HBRs have been used in virtual (i.e., combinations of humans-in-the-loop and computer-based agents) simulations to emulate enemy forces or to supplement friendly forces. However, the distinct advantage of HBRs is in their application to constructive simulations in which human operators are not needed to execute a large number of scenarios, to model operators' reaction and behavior under thousands of combinations of conditions, to assess the resulting performance, and to select an optimal set of variables and constraints. The missile defense studies described next are based on constructive simulations that employ such HBR models.

Over the past 10 years, the air defense and missile defense communities have exhibited a keen interest in studying the affect of automation on crew performance. In 1997, the National Missile Defense (NMD) Joint Program Office (JPO), the GMD project office, the Army Research Laboratory (ARL), the U.S. Army Space Command (ARSPACE), the U.S. Air Force Space Command (AFSPC), Boeing, and TRW initiated a major missile defense operator performance modeling effort. The effort began with the development of a 181-page Operator Task List (OTL). Two separate efforts then attempted to simulate the tasks on this list in the context of realistic scenarios and to validate the simulation. The ARL Human Research and Engineering Directorate (HRED) funded an MA&D effort (1997–2003) to assess battle manager workload. This project aimed to identify the optimal number of battle managers needed to manage a typical BMD battle involving about five ballistic missile threats. In 2001, Boeing tasked TRW to run a similar analysis, and, later, the two independent analyses were compared.

In the 2001 TRW study (September 2001), only the highest level tasks from the OTL (e.g., making a cease engagement or weapons-free decision) were considered. Crews were modeled as teams of operators who were assigned generic roles and worked together to complete the tasks. The model included the time the operators needed to complete each task. These data were obtained in part from ARL and two different Battle Planning Exercises (BPEXs): BPEX 99-1 and BPEX 99-3. Operator and crew performance was measured by calculating their task completion times. Operator stress was modeled by reducing the amount of time required to complete tasks by a fixed percentage (e.g., 20% in one case, 50% in another). The requirement to obtain command approval for decisions increased the time required to complete decision-making tasks by another fixed amount (e.g., 75 seconds). This study considered crews of three, four, and five operators and determined that GMD crews perform best when tasks are distributed among five battle managers. Factors such as the effort needed to manage crew communication as the crew size increased were not taken into account. A more recent study by Aptima, Inc. (Paley, Levchuk, Clark, Miescher, & Baker, 2004) showed that even when crew communication overhead is taken into account, increasing the size of the crew decreases the workload of each crew member. If this finding is true, it suggests that a larger crew size of six, seven, or eight operators might produce even better performance.

In the longer term MA&D study (Walters & Labay, 2003a, 2003b; Walters & Pray, 2003), operators were assigned tasks according to their roles (e.g., battalion director, battle analyst, sensors operator, weapons operator, communications operator). Battle management crew performance was calculated in terms of the number of total tasks the operators could perform, the time to complete the tasks, the number of tasks that were interrupted (and hence dropped), the number that were consequently restarted, and the time that operators spent monitoring the situation. This study was an intense effort to model in detail about 315 of the tasks in the extensive OTL and their low-level task contingencies. MA&D studied the performance of crews made up of 4, 5, and 6 operators during 2002–2003, and the results confirmed the findings of the earlier 2001 TRW study. The MA&D study showed that a five-person crew of battle managers completed a greater number of tasks overall in a shorter time period, dropped fewer tasks because of interruption, restarted a greater number of tasks that were dropped, and spent more time monitoring the situation and gaining situational awareness. One of the main lessons learned was the difficulty in obtaining accurate task times for the tasks in the OTL, in particular those inherently cognitive decision-making tasks.

Decision-making tasks that are inherently cognitive contain a large amount of variability across battle managers.

More recently, the U.S. Air Force Electronic System Command funded an effort to determine the optimal operator task loading and crew configuration (e.g., who should do what, when, where) to conduct a Battle Management, Command and Control (BMC2) mission using the E-10 Multi-Sensor Command and Control Aircraft (MC2A). The E-10A MC2A aircraft supports battle management, intelligence, surveillance, reconnaissance, and selected information warfare functions (Levchuk, Chopra, Paley, Levchuk, & Clark, 2005; Moore, 2004). This study is particularly relevant because of its application to battle management for air and cruise missile defense.

For this effort, Aptima, Inc. (Levchuk et al., 2005; Paley et al., 2004), under contract to the Massachusetts Institute of Technology (MIT) Lincoln Laboratory, developed the Team Optimal Design (TOD) model. First, they created a model that described 33 functions (e.g., process indications and warnings, provide threat updates, determine weapon-to-target pairing) that battle managers carry out while operating the MC2A. These functions were derived from SME working groups,[6] system documentation, and mission scenarios from a Virtual Flag training exercise (Paley et al., 2004). Each function was decomposed into a task flow diagram that described the sequence of tasks required to fulfill the function's goals. For example, the function "assess active threats" involved

---

6 The working groups met on six occasions at Langley Air Force Base (AFB). About 12 U.S. Air Force active duty and civilian SMEs attended each meeting, and a core compliment of about 5 SMEs attended all 6 meetings (personal communication with M. Paley, Aptima, Inc.).

sequences of tasks such as “perform risk assessment for friendly assets” or “identify radar track 28.” The 147 tasks in the task flow diagrams were also assigned attributes such as duration, workload, and information requirements. A series of 54 mission events were then created (e.g., TBM Launch, Red EW Radars Active, Red Strike package ingress). These mission events entailed the execution of the already defined functions.[7] One representative mission required a 25-person MC2A crew to complete 12,246 tasks during a 6-hour period.

Tasks with similar characteristics were grouped into representative task classes. The TOD model considered these classes of tasks, the resources available, and the characteristics of the battle managers (e.g., competence, experience, memory, and learning) to compute the most efficient combination of battle manager roles and responsibilities for a given scenario. Although the computation of optimal crew configurations may not be relevant to our planned OITL experiment, the task, workload, and accuracy models that Aptima, Inc. developed to reach this endpoint can similarly be applied to study the temporal dynamics of battle manager performance and workload.

The TOD model was designed to compute the workload for a battle manager at time t as a function of the classes of all the tasks that the battle manager is performing and the residual workload from previous tasks (which fades over time). Accuracy for a battle manager at time t was calculated as a function of the battle manager’s competence (which is determined by learning rate, memory, and training experience) and workload at that time. For situations in which workload is low, one can choose to model accuracy as high (when the battle manager performs the task automatically) or low (when the battle manager is bored).

Changes in the degree and type of automation changes the crew composition requirements, which, in turn, changes the crew members’ roles and overall performance.

The Aptima E-10A researchers explained that an optimal crew configuration might not exist. Distributing all the necessary tasks to some number of crew members so that none of the battle managers is overloaded might not be possible. In that case, battle managers may end up with overlapping responsibilities, which would increase the need to communicate and coordinate. The overhead of this communication and coordination then factors back into the calculation of overall workload. One of the most important

7 The assignment of specific functions to scenario events was determined by both active duty and civilian SMEs from a number of military and DoD organizations.

lessons learned involved the degree of system automation. Changes in the degree and type of automation would change the crew composition requirements. In turn, the results of the simulation showed that changing the composition of the crew and the crew members' associated roles had the greatest effect on the overall performance (Paley et al., 2004).

D. HOW AUTOMATION CAN HINDER OPERATOR PERFORMANCE

This section reviews literature related to Question 3:

Under what circumstances might automation decrease operator performance and situational awareness while still optimizing operator workload?

In this section, we review research that suggests that an ABMA may, in some cases, decrease operator performance. Although the intent of the ABMA is to decrease operator workload, it may increase the overall level of cognitive effort required by the operator. In addition to requiring the operator to continue to assess the situation and formulate his own decisions, the automated system would require the operator to evaluate the system's recommendations and compare them to his decisions (Hilburn, 2004; Miller & Parasuraman, 2007). These requirements may also lead to additional job preparation and training in "managing the automated battle manager" (Hawley, Mares, & Giammanco, 2006).

Automated decision aids can also produce automation bias, a condition in which operators learn to rely on the cues presented by the automated system as a replacement for their own cognitive effort, human information seeking, and processing (Mosier, Stitka, Heers, & Burdick, 1998). While reliance on these automated decision aids can improve performance by freeing the operator from attending to mundane tasks and enabling him to concentrate on complex cognitive tasks, overreliance results in accidents, especially when the automated system fails (Sheridan & Parasuraman, 2006).

Automation Bias, Complacency, and Supervisory Control Effects

From the late 1970s though the 1980s, industrial engineers, human factors researchers, and psychologists were concerned about the way that human information processing errors were being blamed for several devastating system failures (e.g., the meltdown at Three Mile Island in 1979, the Korean Airlines plane shot down by Soviet fighters in 1983, the USAir B-737 crash in 1989). This concern led to an abundance of research that showed how operator awareness could be reduced to unsafe levels when the human is removed from the control loop and an automated computer controller is responsible for operating the system (Kaber & Endsley, 1997). For example, Wickens (1992) showed that operators respond more slowly to systems that are running in an automated mode. Automation can hamper the development and maintenance of skills required during normal manual operations and increase the time required to train these skills. The time available for training must be distributed across courses for training fundamental skills and courses for training operators how to manage the automated system (Hawley et al., 2006).

Critical operational errors can result from a misallocation of functions between the automated system and the human operators.

Critical operational errors can also result from a misallocation of appropriate functions between the automated system and the human operators (Wickens, 1992). In situations where finding enough skilled operators is difficult and assigning tasks to an automated system may be more cost effective, operators may be reduced to supervisory roles. This situation places them out of the control loop and makes them susceptible to attention-degradation effects.

Examples of automation bias effects in AMD operations are not uncommon. During Operation Iraqi Freedom (OIF), operators' overreliance on a fallible automated system led to two separate fratricide incidents and the loss of three flight crew members (Hawley, 2007). The automated system was the Army's Patriot missile defense system, which had experienced misclassification errors during operational tests before the incidents. The operators' performance was driven by battle management training on rote drills, tactics, techniques, and procedures. Decision-making for cases of track misidentification or misclassification was not comprehensively covered during this training. In the first incident, a British Tornado was misclassified as an antiradiation missile. In the second incident, a Navy F/A-18 was misclassified as a tactical ballistic missile (TBM). Both targets were engaged and destroyed.

An abundance of research directed by the Federal Aviation Administration (FAA) through the 1980s and 1990s studied the phenomena of automation bias, complacency, and other similar issues in air traffic control. For example, Endsley and Rodgers (1996) studied the way in which air traffic controllers distributed their attention among aircraft while observing 15 different scenarios that contained operational errors. They used the Situation Assessment Through Re-creation of Incidents (SATORI) system to simulate the data from actual recorded air traffic control situations and synchronized the simulation with audio tapes of the controller-pilot communication. Occasionally, the researchers froze the simulation, blanked out the screen, and asked the controllers a number of questions. The study showed that the controllers reported only about 67% of the aircraft present on the display and did not generally retain detailed aircraft information (e.g., call signs, groundspeed, and direction). This low level of situational awareness may be explained or intensified by supervisory control effects (described next).

In Endsley and Rodgers' study, the air traffic controllers did not interact with the simulation. Instead, they passively monitored the system. The activities involved in passively monitoring an air traffic control simulation may be similar to the monitoring activities involved in a highly automated environment with an ABMA that provides the maximum level of automation. In this passive mode, the operator may not achieve the same level of attention and situational awareness as when he actively monitors and controls the system. Thus, while automation may reduce operator workload, it may also have the side effect of decreasing operator activity, engagement, and attention. When this happens, operator situational awareness and performance may also decrease. If this is true, it may explain the relatively low level of situational awareness observed in Endsley and Rodgers' study.

While automation may reduce operator workload, it may also decrease operator activity, engagement, and attention, which could lead to a decrease in situational awareness and performance.

This effect can be exacerbated in the presence of novice operators who do not have the tactical and technical knowledge needed to understand the system decision processes. Research by the U.S. Navy suggests that a battle manager's level of experience and his tactical and technical expertise are directly related to his ability to maintain the situational awareness needed to supervise a fully automated system effectively (Hawley & Mares, 2006). For our planned OITL experiments, this research suggests that a novice operator's performance may degrade faster than an expert operator's performance under the fully automated ABMA condition (Level 8 in Table 3).

Kaber and Endsley (1997) conducted a study that examined the specific combinations of human operator and automated system coordination that increase (or decrease) overall system performance. It drew upon a taxonomy that described 10 graded levels of automation from strict manual control to fully automated (see Table 3 for a condensed version).

During the experiment, subjects were asked to eliminate simulated targets that were moving toward the center of the screen. In some cases, the human operator performed the functions of monitoring the system status, generating strategies for eliminating targets, selecting a particular strategy, and implementing this strategy. In other cases, the system performed various combinations of these functions. Kaber and Endsley found that overall performance degraded under the strictly manual control condition and all other conditions in which some of the tasks were automated but that the human was ultimately tasked with implementing the plan. When the level of automation varied across time, subjects had difficulty recovering from situations in which automation included advanced queuing of targets. They became accustomed to focusing on future tasks and tended to neglect present state incidents.

This experiment examined combinations of human operator and automated system functions and confirmed the importance of establishing the appropriate allocation and coordination between these functions. Neither this experiment nor the earlier one (Endsley & Rodgers, 1996) studied the dynamics of how air traffic controllers' attention changes as the number of aircraft in the simulation increases or decreases. We expect this to be a focal point of the OITL AMD experiment.

Cummings and Mitchell (2006) studied the workload and performance of 12 operators as they supervised 4 simulated unmanned aerial vehicles (UAVs) that were tasked to destroy a set of time-sensitive targets. (Nine of the 12 participants were active duty United States Air Force (USAF) officers or Reserve Officer Training Corps (ROTC) students.) The operators were responsible for tasks such as assigning or unassigning targets to UAV mission plans, arming and firing payloads, and ordering UAVs to return to base. Three conditions representing different levels of automated decision support were tested:

- The first level involved a manual decision-aid display containing a series of visual timelines showing the scheduling of ATO events associated with each UAV (see Figure 6).
- The second level included an automated decision aid in the form of the visual timelines alongside a series of computer-based recommendations that the operator could accept or reject.
- The third level was a fully automated management-by-exception system that executed arming and firing actions according to the rules of engagement.

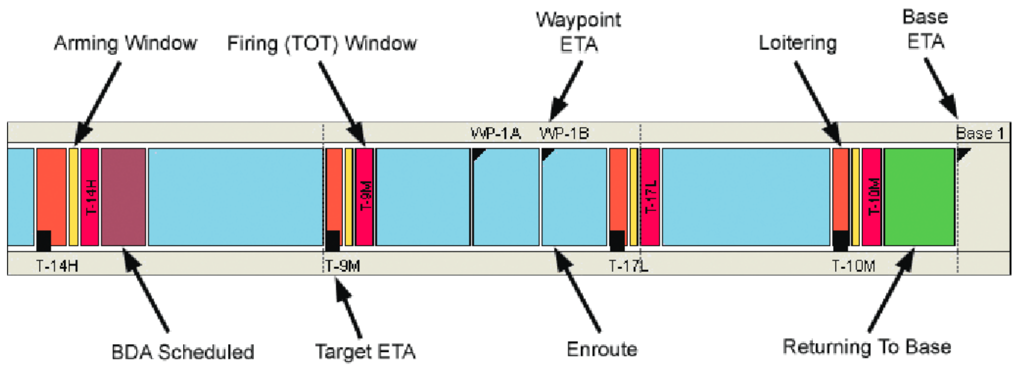

Figure 6. Visual timeline decision aid for managing and scheduling the UAV ATO (from Cummings & Mitchell, 2006)

Each condition progressively automated more functions involved in the management and scheduling of the ATO. Operators using the fully automated system had an opportunity to intervene and veto each action 30 seconds before it occurred. The scenarios were scripted according to two levels of difficulty (high replanning and low replanning) depending on the frequency of replanning that was required in the scenario to address emergent threats, new tasking from superiors, and system failures.

The results of Cummings and Mitchell's study showed that in the high replanning condition, the operators who used the automated timeline decision aid had lower performance scores and higher subjective workload scores than those who used the manual timeline and those who used the fully automated management-by-exception system. In fact, the automated decision aid produced the poorest scores overall and the lowest situational awareness (measured by the operators' subjective assessment of their comprehension of the current situation). These results suggest that even an arbitrary and conservative level of automation does not necessarily improve performance under high-workload conditions. This study also illustrates the complexity of the interaction among the human and automation-related variables that affect the resulting workload and performance of the human-computer partnership.

One such complex interaction involves the human operator's perception and trust of the system's recommendations. Skitka, Mosier, and Burdick (1999) point out the extensive research in social psychology showing that a person is likely to harm others if directed to by an authority figure. To the extent that a person perceives an automated decision aid as having authority, there is reason to believe that the person might similarly follow the system's recommendations without further consideration. Stitka et al. (1999) tested this hypothesis by asking 80 participants to perform a set of tasks designed to simulate the monitoring and tracking of commercial airlines. In the experimental condition, the participants had access to an automated monitoring aid that prompted them about various system events. In the control condition, the participants used the same system but did not have access to the decision aid. The participants were specifically told that the automated decision aid was not always accurate and that the other gauges and instruments (available in both conditions) were always 100% accurate. The results of this experiment confirmed Stitka et al.'s hypothesis. Not only did the participants in the experimental condition follow the advice of the decision aid when the other instruments provided contradictory evidence, but they were also less vigilant than the control condition, missing a significantly larger number of events that occurred without a system prompt.

To the extent that a person perceives an automated decision aid as having authority, he will follow the system's recommendations even in the face of contradictory evidence.

E. SUMMARY

This report summarizes the findings from an array of literature related to how an AMD battle manager's performance degrades with increased workload in the context of various realistic scenarios. We also discussed how an ABMA can moderate this degradation. The reviewed literature was organized according to three research questions:

1. Without automation assistance, how many decisions can an operator handle per unit time? At what point does operator performance drop off, and does it drop off gradually or abruptly?
2. Under what circumstances will automation improve operator performance and optimize operator workload?
3. Under what circumstances might automation decrease operator performance and situational awareness while still optimizing operator workload?

For the first question, without the assistance of automation, a battle manager's performance will degrade as the complexity of the task increases. If human memory capacity is limited in the way the psychological and air defense research suggests, we should expect the performance of an AMD battle manager to decline rapidly when he becomes overloaded with more than seven entities or decisions. The complexity of the task can be mediated by several factors, including the fidelity and design of the operator display and the level of automation of the system. An operator can improve his performance by increasing his cognitive capacity, restructuring his knowledge, and gaining experience in the domain. Other factors, such as an operator's risk-taking and cultural behaviors, can also affect his performance.

For the second question, one of the prominent factors that affects an operator's performance is the level of automation. Section C.1 outlined four different stages and eight different levels at which automation can enhance system and human performance. One of the studies reviewed (Bruni & Cummings, 2007; Cummings & Bruni, in press) demonstrated the utility of automation in assisting the human battle manager and the importance of ensuring that he has access to the data needed for decision-making, even when the automated system will be recommending or making pairing decisions. At the same time, this study indicated that providing an extensive amount of drill-down information in a time-sensitive situation would compel the battle manager to review all the data, increasing the problem-solving time. Section C.2 explained how changes in the degree and type of automation introduced into the system would change the crew composition requirements. In turn, changing the composition of the crew and the crew members' associated roles affected their overall performance.

For the third question, an abundance of research indicates that while automation may decrease operator workload, it may also, paradoxically, increase the overall level of cognitive effort required by the operator. In addition to requiring operators to continue to assess the situation and formulate their own decisions, automated systems require operators to evaluate the system's recommendations and compare these recommendations with their own. This additional cognitive effort in "managing the automated battle manager" is also prone to the consequences of automation bias, a situation in which operators learn to trust and rely on the cues that are presented by the automated system as a replacement for their cognitive effort, human information seeking, and processing. There is no shortage of research showing how overreliance on automation results in fatal accidents when the automated system fails.

Some of the studies described in this report have direct implications for the design and analyses of our OITL experiments and the design of the battle manager simulation interface. Because different battle managers are likely to experience different workloads at different times for the same scenario, the OITL experiment participants should include battle managers who have a range of abilities (e.g., novice, intermediate, expert) and backgrounds. The simulation should alert the battle manager about critical situations, and the design of the interface should not impede the battle manager's ability to access the decision-making data. To the extent possible, the simulation environment should encourage a trusting human-system partnership, but it should not induce a false perception of trust if the decision-making data are not reliable. The reliability of the decision-making data should be apparent to the battle manager and should play a major role in assessing the most appropriate level of automation. A battle manager's performance is expected to degrade as the complexity (e.g., the timing, quantity and order of events, and degree of uncertainty) of the task increases. As the battle manager's performance degrades, the most appropriate level of automation should be based on an understanding of human behavioral strengths, tendencies, and vulnerabilities and on the consequences of making mistakes. If automation decreases operator activity, engagement, and attention and leads to a decrease in situational awareness and performance, battle manager performance should decrease at the highest ABMA level. The OITL experiments should be designed to test these hypotheses, and the selected analysis methods should address these considerations.

REFERENCES

Adelman, L., Bresnick, T., Christian, M., Gualtieri, J., & Minionis, D. (1997). Demonstrating the effect of context on order effects for an army air defense task using the patriot simulator. *Journal of Behavioral Decision Making*, *10*(4), 327–342.

Adelman, L., Miller, S. L., & Yeo, C. (2004). Testing the effectiveness of icons for supporting distributed team decision making under time pressure. *IEEE Transactions on Systems, Man and Cybernetics—Part A: Systems and Humans, 34*(2), 179–189.

Anderson, J. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369–406.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Bruni, S., & Cummings, M. L. (2007). Determining operators' cognitive strategies in a mission planning resource allocation task. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* (under review).

Cannon-Bowers, J., & Salas, E. (1998). Individual and team decision-making under stress: Theoretical underpinnings. In J. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress* (pp. 17–38). Washington D.C.: American Psychological Association.

Cummings, M. L., & Bruni, S. (in press). Collaborative human-computer decision-making in network centric warfare. *International Journal of Aviation Psychology*.

Cummings, M. L., & Mitchell, P. (2006). Automated scheduling decision support for supervisory control of multiple UAVs. *AIAA Journal of Aerospace Computing, Information, and Communication, 3*(6), 294–308.

Cummings, M. L., & Tsonis, C. G. (2006). Partitioning complexity in air traffic management tasks. *International Journal of Aviation Psychology, 16*(3), 277–295.

Endsley, M., & Rodgers, M. (1996). Attention distribution and situation awareness in air traffic control. *Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society*, 82–85.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23.

Gott, S., & Lesgold, A. (2000). Competence in the workplace: How cognitive performance models and situated instruction can accelerate skill acquisition. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 5). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 239–250). Amsterdam: North-Holland Press.

Hawley, J. (2007). *Job preparation for the 21st century warfighter: Lessons from Patriot after Operation Iraqi Freedom*. Paper presented at the Human Systems Integration Symposium.

Hawley, J., & Mares, A. (2006). *Developing effective human supervisory control for air and missile defense systems* (ARL-TR-3742). Fort Bliss, TX: Army Research Laboratory,.

Hawley, J., Mares, A., & Giammanco, C. (2006). *Training for effective human supervisory control of air and missile defense systems* (ARL-TR-3765). Adelphi, MD: Army Research Laboratory.

Hilburn, B. (2004). *Cognitive complexity in air traffic control: A literature review*. European Organisation for the Safety of Air Navigation (EUROCONTROL) Experimental Centre (EEC).

Histon, J. M., & Hansman, R. J. (2002). *The impact of structure on cognitive complexity in air traffic control* (ICAT-2002-4). Cambridge, MA: MIT International Center for Air Transportation.

Hofstede, G. H. (1984). *Culture's consequences: International differences in work-related values*. Beverly Hills: Sage Publications.

Hurst, M., & Rose, R. (1978). Objective job difficulty, behavioral response, and sector characteristics in air route traffic control centres. *Ergonomics, 21*, 697–708.

Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress, 16*(3), 126–131.

Kaempf, G., Wolf, S., & Miller, T. (1993). Decision making in the AEGIS combat information center. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, 1107–1111.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–292.

Klein, H. A., Pongonis, A., & Klein, G. (2000). *Cultural barriers to multinational C2 decision-making.* Paper presented at the Command and Control Research and Technology Symposium, Naval Postgraduate School, Monterey, CA.

Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980). Models of competence in solving physics problems. *Cognitive Science, 4*, 317–345.

Levchuk, G., Chopra, K., Paley, M., Levchuk, Y., & Clark, D. (2005). *Model-based organization manning, strategy, and structure design via Team Optimal Design (TOD) methodology*. Paper presented at the 10th International Command and Control Research and Technology Symposium.

McDermott, P., Klein, G., Thordsen, M., Ransom, S., & Paley, M. (2000). *Representing the cognitive demands of new systems: A decision-centered design approach*: (AFRL-HE-WP-TR-2000-0023). Wright-Patterson AFB, OH: Air Force Research Laboratory, Human Effectiveness Directorate.

Miller, C., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors, 49*(1), 57–75.

Miller, G. (1955). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 101*(2), 343–352.

Mogford, R. H., Guttman, J. A., Morrow, S. L., & Kopardekar, P. (1995). *The complexity construct in air traffic control: A review and synthesis of the literature* (Technical Note DOT/FAA/CT-TN95/22). Atlantic City International Airport: Federal Aviation Administration William J. Hughes Technical Center.

Mogford, R. H., Murphy, E. D., & Guttman, J. A. (1994). Using knowledge exploration tools to study airspace complexity in air traffic control. *International Journal of Aviation Psychology, 4*(1), 29.

Moore, J. (2004). *E-10A multi-sensor command and control aircraft:* (NAIC-1574-3472-04). Wright-Patterson AFB, OH: National Air Intelligence Center.

Morrison, J. (2003). *A review of computer-based human behavior representations and their relation to military simulations* (IDA Paper P-3845). Alexandria, VA: Institute for Defense Analyses.

Mosier, K., Stitka, L., Heers, S., & Burdick, M. (1998). Automation bias: Decision-making and performance in high-tech cockpits. *The International Journal of Aviation Psychology, 8*(1), 47–63.

Mui, R., Bagnall, T., LaVine, N., Peters, S., & Sargent, B (2004). *Cultural modeling of command and control echelons*: (AFRL-HE-WP-TR-2004-0112). Wright-Patterson AFB, OH: Air Force Research Laboratory, Human Effectiveness Directorate.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Lawrence Erlbaum Associates.

Paley, M., Levchuk, G., Clark, D., Miescher, S., & Baker, K. (2004). *Modeling, simulation, and experimentation for the multi-sensor command and control aircraft (MC2A): Final report* (AP-R-1231). Prepared by Aptima, Inc. for MIT/Lincoln Laboratory.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 185–218). Amsterdam: Elsevier.

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology: An International Review, 53*(1), 61–86.

Schafer Corporation. (15 January 2005). *Human engineered intelligent agent phase II SBIR quarterly report* (Contract DASG60-03-C-0078). SBIR Phase II Briefing to the GMB Program Management Office (GFC Products Division).

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. Cambridge, MA: MIT Press.

Sheridan, T. B., & Parasuraman, R. (2006). Human-automation interaction. *Reviews of Human Factors and Ergonomics, 1*, 89–129.

Sheridan, T. B., & Verplank, W. (1978). *Human and computer control of undersea teleoperators*. Cambridge, MA: MIT, Department of Mechanical Engineering, Man-Machine Systems Laboratory.

Simon, H. (1974). How big is a chunk? *Science, 183*(4124), 482–488.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies, 51*(5), 991–1006.

Smith, C., Johnston, J., & Paris, C. (2004). Decision support for air warfare: Detection of deceptive threats. *Group Decision and Negotiation, 13*(2), 129–148.

TRW. (September 2001). *Command and control center personnel requirements analysis*. Huntsville, AL: GMD GFC Program Office Critical Design Review.

Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics, 39*(3), 358–381.

Walters, B., & Labay, S. (2003a). *The dynamic Ground-based Midcourse Defense Fire Control (GFC) operator model: Supporting operator training and facilities planning*. Final Report prepared for the U.S. Army Research Laboratory, Contract No. DAAD19-01-C-0065. Boulder, CO: Micro Analysis and Design, Inc.

Walters, B., & Labay, S. (2003b). *Validation of the dynamic Ground-based Midcourse Defense Model*. Final Report prepared for the U.S. Army Research Laboratory, Contract No. DAAD17-00-A-5003. Boulder, CO: Micro Analysis and Design, Inc.

Walters, B., & Pray, S. (2003). *Expansion and validation of the Ground-based Midcourse Defense Dynamic Operator Task List Model*. Final Report prepared for the U.S. Army Research Laboratory, Contract No. DAAD17-00-A-5003. Boulder, CO: Micro Analysis and Design, Inc.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: Harper Collins.

Wickens, C. D., Mavor, A. S., Parasuraman, R., & McGee, J. P. (1998). *The future of air traffic control: Human operators and automation*. Washington, D.C.: National Academy Press.

GLOSSARY

AAR	After Action Report
ABL	Airborne Laser
ABMA	automated battle management aid
AFB	Air Force Base
AFRL	Air Force Research Laboratory
AFSPC	U.S. Air Force Space Command
AMD	air and missile defense
ARL	Army Research Laboratory
ARSPACE	U.S. Army Space Command
ATO	Air Tasking Order
AWACS	Airborne Warning and Control System
BDA	Battle Damage Assessment
BMC2	Battle Management, Command and Control
BPEX	Battle Planning Exercise
C2	command and control
DSMAC	Digital Scene Matching Area Correlation
ETA	Estimated Time of Arrival
EW	early warning
FAA	Federal Aviation Administration
GFC	GMD Fire Control (GFC)
GMD	Ground-Based Midcourse Defense
GPS	Global Positioning System
HBR	Human Behavior Representation
HMI	human-machine interface
HRED	Human Research and Engineering Directorate
IADS	Integrated Air Defense System
IDA	Institute for Defense Analyses

JEFX	Joint Expeditionary Force Exercise
JLENS	Joint Land Attack Cruise Missile Defense (LACMD) Elevated Netted Sensor System (JLENS)
JPO	Joint Program Office
JTAMDO	Joint Theater Air and Missile Defense Organization
LACMD	Land Attack Cruise Missile Defense
MA&D	Micro Analysis and Design, Inc
MC2A	Multi-Sensor Command and Control Aircraft
MIT	Massachusetts Institute of Technology
NAS	National Academy of Sciences
NASA	National Aeronautics and Space Administration
NGC	Northrop Grumman Corporation
NMD	National Missile Defense
OIF	Operation Iraqi Freedom
OITL	operator-in-the-loop
OTL	Operator Task List
POC	point of contact
PM	program manager
ROTC	Reserve Officer Training Corps
RV	reentry vehicle
SATORI	Situation Assessment Through Re-creation of Incidents
SBIR	Small Business Innovative Research
SIAP	single integrated air picture
SME	subject matter expert
SOP	standard operating procedure
SWAT	Subjective Workload Assessment Technique
TBM	tactical ballistic missile
THAAD	Theater High Altitude Area Defense
TLAM	Tomahawk Land Attack Missile
TLX	Task Load Index
TOD	Team Optimal Design
TOT	Time on Target

VWC	Virtual Warfare Center
UAV	unmanned aerial vehicle
USAF	United States Air Force
USS	United States Ship
WM	working memory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED *(From–To)*
January 2008	Final	June 2007 – January 2008

4. TITLE AND SUBTITLE

The Effects of Automation on Battle Manager Workload and Performance

5a. CONTRACT NUMBER
DASW01-04-C-0003

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Amy Soller
John Morrison

5d. PROJECT NUMBER

5e. TASK NUMBER
CA-2-1617.01

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311-1882

8. PERFORMING ORGANIZATION REPORT NUMBER

IDA Document D-3523

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

JTAMDO
1851 S. Bell Street
Crystal Mall 3, Suite 511
Arlington, VA 22202

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited. (19 May 2008)

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report summarizes the literature reviewed in preparation for planning and executing a series of controlled, operator-in-the-loop (OITL) experiments to determine how an air and missile defense (AMD) battle manager's performance degrades with increased workload and how automated battle management aids (ABMA) can moderate this degradation. The sources for this survey range from studies that describe the basic limits of human memory capacity to those that assess the number of battle managers needed to operate a partially automated missile defense system. The research indicates that without the assistance of automation, a battle manager's performance will degrade as the complexity of the task increases; however, human performance can vary considerably across experience levels and tasks. This report outlines four different stages and eight different levels at which automation can enhance system and human performance. An abundance of research indicates that while automation may decrease operator workload, it may also decrease operator activity, engagement, and attention, leading to a decrease in situational awareness and performance.

15. SUBJECT TERMS

air and missile defense, automation, battle management, human factors, decision-aids, decision-making, workload

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON LCDR David Weller
a. REPORT Uncl.	b. ABSTRACT Uncl.	c. THIS PAGE Uncl.	SAR	52	19b. TELEPHONE NUMBER *(include area code)* 703-602-5123

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18